Sentiment Analysis in the Age of Generative AI

Abstract

In the rapidly advancing age of Generative AI, Large Language Models (LLMs) such as ChatGPT stand at the forefront of disrupting marketing practice and research. This paper presents a comprehensive exploration of LLMs' proficiency in sentiment analysis, a core task in marketing research for understanding consumer emotions, opinions, and perceptions. We benchmark the performance of three state-of-the-art LLMs, i.e., GPT-3.5, GPT-4, and Llama 2, against established, high-performing transfer learning models. Despite their zero-shot nature, our research reveals that LLMs can not only compete with but in some cases also surpass traditional transfer learning methods in terms of sentiment classification accuracy. We investigate the influence of textual data characteristics and analytical procedures on classification accuracy, shedding light on how data origin, text complexity, and prompting techniques impact LLM performance. We find that linguistic features such as the presence of lengthy, content-laden words improve classification performance, while other features such as single-sentence reviews and less structured social media text documents reduce performance. Further, we explore the explainability of sentiment classifications generated by LLMs. The findings indicate that LLMs, especially Llama 2, offer remarkable classification explanations, highlighting their advanced human-like reasoning capabilities. Collectively, this paper enriches the current understanding of sentiment analysis, providing valuable insights and guidance for the selection of suitable methods by marketing researchers and practitioners in the age of Generative AI.